

Request for Judicial Notice

EXHIBIT 4

Date of Hearing: April 27, 2021

ASSEMBLY COMMITTEE ON JUDICIARY
Mark Stone, Chair
AB 587 (Gabriel) – As Amended March 25, 2021

As Proposed to be Amended

SUBJECT: SOCIAL MEDIA COMPANIES: TERMS OF SERVICE

KEY ISSUES:

- 1) SHOULD A SOCIAL MEDIA COMPANY BE REQUIRED TO POST TERMS OF SERVICE, ESPECIALLY AS TO THE ONLINE BEHAVIOR AND ACTIVITIES THAT IT EITHER PERMITS OR PROHIBITS?
- 2) SHOULD A SOCIAL MEDIA COMPANY BE REQUIRED TO SUBMIT QUARTERLY REPORTS DESCRIBING ITS CONTENT MODERATION POLICIES TO THE ATTORNEY GENERAL, AND SHOULD THE ATTORNEY GENERAL BE REQUIRED TO POST THOSE REPORTS ON ITS WEBSITE?

SYNOPSIS

This bill is one of three that the Committee had heard this year dealing with a social media company's ability to monitor and moderate posted content and users' online conduct. AB 35 (Chau) would require social media platforms to post what, if anything, they do to stop the spread of misinformation and violent or hateful content. AB 1114 (Gallagher) would require social media platforms to disclose how they address "unprotected" speech (obscenity, true threats, etc.); but for everything else, in its original form, AB 1114 would have made social media platforms a traditional public forum, thereby limiting their ability to remove objectionable content or deny access to certain users. The bill now before the Committee seeks transparency by requiring social media companies to post their "terms of service," especially as to the kinds of online behavior and activities they permit and, conversely, the kinds of behavior and activities that will lead the social media company to ban the user from its service. In addition, this bill would require social media companies to submit quarterly reports to the California Attorney General. In addition to including a copy of the company's terms of service, the reports would contain detailed descriptions of the methods used to moderate online activity. The report would also contain quantitative information, such as the number of times the social media company took action against users or content. Although the bill requires the Attorney General to post these reports on its website, it is not clear what else, if anything, the Attorney General would be expected to do with the information in the reports.

The bill is supported by several civil rights groups and ethnic and religious associations, who contend that social media has become a powerful and largely unregulated platform for groups espousing hate, violence, and bigotry. The bill is opposed by social media companies who fear that the information disclosed will allow "bad actors" to circumvent moderation control; they also believe the quarterly reports to the Attorney General are both burdensome and unnecessary. The bill passed out of the Assembly Privacy and Consumer Protection Committee on a 9-0 vote with an understanding that the author would take amendments in this Committee. Those

amendments are included in the summary below. However, the amendments have not removed the opposition.

SUMMARY: Requires a social media company to post terms of service (TOS) related to permitted and prohibited user behavior and activity and requires the social media company to submit reports and other relevant documents to the Attorney General (AG), as specified.

Specifically, **this bill:**

- 1) Requires a social media company to post their TOS in a manner reasonably designed to inform all users of the social media company's service of the existence and contents of the TOS, and require the TOS to include all of the following:
 - a) Contact information for the purpose of allowing users to ask the social media company questions about the terms of service.
 - b) A description of the process that users must follow to flag content, groups, or other users that they believe violate the terms of service, and the social media company's commitments on response and resolution time.
 - c) A list of potential actions the social media company may take against any item of content, or a user, or group of users, including, but not limited to, removal, demonetization, de-prioritization, or banning.
- 2) Requires the TOS to be available in all languages in which the social media company offers product features, including but not limited to menus and prompts.
- 3) Provides that a social media company is in violation of the above provisions if it fails to comply with the provisions of this section within 30 days of being notified of noncompliance by the AG.
- 4) Requires a social media company to submit to the AG, on a quarterly basis, a TOS report, covering activity within the previous three months. The AG shall post on its website all reports submitted pursuant to this requirement. The first report shall be submitted no later than July 1, 2022, and shall include:
 - a) A complete and detailed description of any changes to the TOS since the last report;
 - b) A statement of whether the current version of the TOS defines each of the following categories, including their definitions, if applicable: (1) hate speech or racism; (2) extremism or radicalization; (3) disinformation or misinformation; (4) harassment; or (5) foreign political interference;
 - c) A complete and detailed description of content moderation practices used by the social media company, including, but not limited to: (1) any existing policies intended to address the categories of content described immediately above; (2) any rules or guidelines regarding automated content moderation systems, as specified; (3) any training materials provided to content moderators, including educational materials; (4) responses to user reports of violations of the terms; (5) any rules, guidelines, product changes and content moderator training materials that cover how the social media company would remove individual pieces of content, users, or groups that violate the terms, or take

broader action against individual users or against groups of users that violate the terms; and (6) the languages in which the social media company offers product features, as specified;

- d) Information, de-identified and disaggregated, as specified, on content that was flagged by the social media company as content belonging to any of the categories described above, including all of the following: (1) the total number of flagged items of content; (2) the total number of actioned items of content, including the total number of actioned items of content that were removed, demonetized, or deprioritized; (3) the number of times actioned items of content were viewed by users, shared, and the number of users that viewed that content before it was actioned; (4) the number of times users appealed social media company actions and reversals of those actions on appeal, as specified.
- 5) Provides that any violation of the above provisions shall be actionable under the Unfair Competition Law in addition to any other applicable state or federal law.
- 6) Provides definitions including:
 - a) “Actioned” means a social media company, due to a suspected or confirmed violation of the TOS, has taken some form of disciplinary action, as specified.
 - b) “Content” means media, including, but not limited to, text, images, videos, and groups of users that are created, posted, shared, or otherwise interacted with by users on an internet-based service.
 - c) “Social media company” means a person or entity that owns or operates a public-facing internet-based service that generated at least \$100,000,000 in gross revenue during the preceding year, and that allows users in the State to do all of the following: (1) construct a public or semipublic profile within a bounded system created by the service; (2) populate a list of other users with whom an individual shares a connection within the system; and (3) view and navigate a list of the individual’s connections and the connections made by other individuals within the system.
- 7) States that it is the intent of the Legislature that a social media company that violates this chapter shall be subject to meaningful remedies sufficient to induce compliance with the provisions above.

EXISTING LAW:

- 1) Provides, under the Communications Decency Act of 1996, that “no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider,” and affords broad protection from civil liability for the good faith content moderation decisions of interactive computer services. (47 U.S.C. Section 230(c)(1) and (2).)
- 2) Provides, under the Unfair Competition Law, for specific or preventive relief to enforce a penalty, forfeiture, or penal law in the case of unfair competition; and defines unfair competition to mean any unlawful, unfair or fraudulent business act or practice and unfair, deceptive, and untrue or misleading advertising. (Business & Professions Code Section 17200 *et seq.*)

- 3) Requires an operator of a commercial website or online service that collects personally identifiable information, as specified, to make its privacy policy available to consumers, as specified. (Civil Code Section 1798.130.)

FISCAL EFFECT: As currently in print this bill is keyed fiscal.

COMMENTS: According to the author:

Californians are becoming increasingly alarmed about the role of social media in promoting hate, disinformation, conspiracy theories, and extreme political polarization. Despite the widespread nature of these concerns, efforts by social media companies to self-police such content have been widely criticized as opaque, arbitrary, biased, and inadequate. While some platforms share limited information about their efforts, the current lack of transparency has exacerbated concerns about the intent, enforcement, and impact of corporate policies, and deprived policymakers and the general public of critical data and metrics regarding the scope and scale of online hate and disinformation. The public and policymakers deserve to know when and how social media companies are amplifying certain voices and silencing others. This is an important first step in a broader effort to protect our democracy and better regulate social media platforms.

What is the problem that this bill seeks to solve? According to the author and supporters of the bill, social media has become a powerful and largely unregulated platform for groups espousing hate, violence, bigotry, conspiracy theories, and disinformation. For example, the author cites a recent study of Twitter posts from 100 U.S. cities. The study found that the greater proportion of tweets related to race- and ethnicity-based discrimination in a given city, the more hate crimes were occurring in that city. The author also cites the case of Robert Bowers, the man who murdered eleven elderly worshipers at a Pennsylvania synagogue in 2018. Bowers had been active on Gab, a Twitter-like site used by white supremacists. The author contends that other investigations have suggested that the riots at the Capitol in early January of this year were encouraged by posts on social media sites.

Although this measure does not require social media companies to moderate or remove hateful or incendiary content, the author nonetheless sees the bill as “an important first step” in protecting our democracy from the dangerously divisive content that has become all too common on social media. In essence, AB 587 is a transparency measure. It would require social media companies to post their “terms of service,” especially as to the kinds of online behavior and activities they permit and, conversely, what kinds of behavior and activities subject a user to a temporary or permanent prohibition from using the social media service. Presumably, requiring these postings will serve three purposes. First, it will let users know what social media platforms do to flag and remove certain kinds of content, which may affect what sites users prefer to use. Second, it lets users know in advance what kind of content or conduct could lead to their being temporarily or permanently banned from using the social media service. Third, if social media companies are forced to disclose what they do in this regard, it may pressure them to become better corporate citizens by doing more to eliminate hate speech and disinformation.

Quarterly reports to the Attorney General. In addition to requiring social media companies to post terms of service for the user’s edification, this bill would also require social media companies to submit quarterly reports to the AG. In addition to including a copy of the company’s terms of services, the reports would also contain detailed descriptions of the methods that a company uses to moderate online activity. Specifically, the reports must contain a

“complete and detailed description” of the company’s policies, if any, for addressing hate speech or racism, extremism or radicalization, disinformation or misinformation, harassment, and foreign political interference. The report would also include detailed descriptions on a variety of other topics: the use of automated content moderation systems; the use of any training materials used to educate content moderators; how the company responds to user reports of violations of the terms of service; and any rules, guidelines, or training materials that cover how the company would remove individual pieces of content. The bill would also require the reports to contain quantitative information, including, among other things, the number of times that the social media company flagged or took action against content and the number of times that the company took action against users or groups of users responsible for the content. Finally, the reports would include information on *how* content was handled, including whether action was taken by employees, or by artificial intelligence software. The bill requires the AG to post these reports on its website. Other than positing these reports on its website, it is not clear what else, if anything, the Attorney General would be expected to do with the information in the reports. Although the bill does not yet prescribe a penalty for violations, it states the intent of the Legislature that violations of the bill’s requirements will be “subject to meaningful remedies sufficient to induce compliance,” presumably to be specified in the future.

Opposition concerns. Although some of the amendments taken by the author have assuaged some opposition concerns – such as eliminating a provision allowing the AG to demand documents from the social media companies – the social media companies, along with the California Chamber of Commerce, continue to oppose this measure.

Opposition concerns can be placed into two general categories. First, the opposition fears that the “complete and detailed” information required in the reports to the AG may include information that will allow “bad actors” to circumvent the social media companies content moderation policies – especially because the reports require detailed description on “training materials” and descriptions of the company’s use of automated systems and artificial intelligence. Second, the opposition questions the need for these reports. If the point of the bill is transparency – to inform consumers about the company’s content moderation rules and policies – then that objective is served by requiring the social media company to post its terms of service. Users, the opponents contend, are more likely to check the terms of service of a social media platform than they are to visit the AG’s website. Of course, the proponents of the bill could counter that the reports to the AG will contain *more* information than is required in the terms of service, including more detail on the methods used and numbers of times that the company has taken particular action. Opponents believe, however, that the cost of the reports is not justified by any additional information contained in the reports, which may not be any more useful to the user than the information already available in the company’s terms of service.

In light of the concerns raised by the opposition, the author may wish to consider, in particular, whether some of the information required in the reports – such as training manuals and descriptions of automated moderation systems – could, in fact, be exploited by those who want to circumvent a company’s moderation policies. In addition, the author may wish to consider the extent to which the information required in the terms of service mirrors the information required in the quarterly reports. If the information is not substantially different, then it is difficult to see what purpose is served by requiring the submission of the reports and their posting on a public website. Does the information provided in the AG reports provide information that is more useful to the user than that provided in the terms of service?

ARGUMENTS IN SUPPORT: This bill is supported by a wide coalition of groups who are particularly concerned about increased use of social media to promote hate, and sometimes incite violence, against groups based upon race, ethnicity, religion, or gender identity, among other factors. For example, the Anti-Defamation League (ADL), who writes on behalf of itself and several other organizations, supports AB 587 because it will “require social media platforms to publicly disclose their content moderation policies regarding online hate/racism, disinformation, extremism, and harassment, as well as regularly publish standardized key metrics and data around the enforcement of those policies.” In addition to citing studies suggesting an increase of online hate and extremism, ADL contends that “efforts by social media companies to self-police such content have been opaque, arbitrary, biased, and inadequate.” ADL believes that “AB 587 would address this troubling lack of transparency by requiring social media platforms to publicly disclose their corporate policies and report key data and metrics around the enforcement of their policies.”

Several other civil rights groups and ethnic and religious associations support this bill for substantially the same reasons as those articulated by the ADL.

ARGUMENTS IN OPPOSITION: Opponents of the measure, including the California Chambers of Commerce and associations representing internet-based and social media companies, claim that the disclosures required by this bill will expose social media moderation practices to “exploitation by bad actors.” In addition, opponents claim that the reporting requirements in the bill are “burdensome” and “unworkable.” While the opponents applaud the author’s “laudable goal of increasing transparency around content moderation practices,” they contend that the bill in print could ultimately “frustrate this goal.”

First, opponents contend that exposing too many details about content moderations practices could be exploited by “bad actors,” who would use that information to circumvent content moderation policies and raise new security threats. “To avoid undermining the goals of the bill and the work our companies have already undertaken to combat harmful content,” the opponents believe that the bill should not require the disclosure of such detailed information, especially the information on training materials, the automated content moderation systems used to enforce terms of service, and information on how the social media goes about removing content.

Second, in addition to potentially providing information to bad actors, opponents contend that the reporting requirements are “burdensome” and “will not advance the bill’s purpose.” Although the proposed amendments reduce the reporting requirements somewhat, the opponents contend that requiring “businesses to report detailed metrics on a quarterly basis” would be a nearly impossible task “due to the need to review, analyze, and adjudicate actioned content.” Moreover, “the sheer volume of content our companies review makes it similarly difficult to report data on an individual pieces of content.”

[Note: Other portions of the opposition letter concern provisions that have since been deleted from bill; nonetheless, opponents have informed that Committee that the proposed amendments do not eliminate their opposition, and, indeed, that adding a new provision requiring reports to contain a “complete and detailed description” of certain content moderation practices raises new concerns.]

REGISTERED SUPPORT / OPPOSITION:

Support

AAUW Camarillo Branch
American Jewish Committee - Los Angeles
Anti-defamation League
Armenian Assembly of America
Armenian National Committee of America - Western Region
California Asian Pacific American Bar Association
California League of United Latin American Citizens
Center for the Study of Hate & Extremism - California State University, San Bernardino
Common Sense
Hindu American Foundation, INC.
Islamic Networks INC.
Israeli-American Civic Action Network
Japanese American Citizens League, Berkeley Chapter
Jewish Center for Justice
Jewish Public Affairs Committee
Korean American Bar Association of Northern California
Maplight
National Hispanic Media Coalition
Progressive Zionists of California
Sikh American Legal Defense and Education Fund (SALDEF)
Simon Wiesenthal Center, INC.
Stonewall Democratic Club

Opposition

California Chamber of Commerce
Civil Justice Association of California
Internet Association
MPA the Association of Magazine Media
TechNet

Analysis Prepared by: Thomas Clark / JUD. / (916) 319-2334